



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
-----------------	-------------	----------------------	---------------------	------------------

10/539,047

04/02/2007

Kazuhiro Nakadai

052696

7377

38834

7590

06/24/2011

WESTERMAN, HATTORI, DANIELS & ADRIAN, LLP
1250 CONNECTICUT AVENUE, NW
SUITE 700
WASHINGTON, DC 20036

EXAMINER

EDWARDS, JERRAH

ART UNIT

PAPER NUMBER

3667

NOTIFICATION DATE

DELIVERY MODE

06/24/2011

ELECTRONIC

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Notice of the Office communication was sent electronically on above-indicated "Notification Date" to the following e-mail address(es):

patentmail@whda.com

Office Action Summary	Application No. 10/539,047	Applicant(s) NAKADAI ET AL.	
	Examiner JERRAH EDWARDS	Art Unit 3667	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 19 April 2011.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-13 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-13 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- | | |
|---|---|
| 1) <input checked="" type="checkbox"/> Notice of References Cited (PTO-892) | 4) <input type="checkbox"/> Interview Summary (PTO-413) |
| 2) <input type="checkbox"/> Notice of Draftperson's Patent Drawing Review (PTO-948) | Paper No(s)/Mail Date. _____ |
| 3) <input type="checkbox"/> Information Disclosure Statement(s) (PTO/SB/08) | 5) <input type="checkbox"/> Notice of Informal Patent Application |
| Paper No(s)/Mail Date _____ | 6) <input type="checkbox"/> Other: _____ |

DETAILED ACTION

Allowable Subject Matter

1. The indicated allowability of claims 2-13 is withdrawn in view of the newly discovered reference(s) below. Rejections based on the newly cited reference(s) follow.

Claim Rejections - 35 USC § 103

2. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

3. Claims 1-2, 4-9 and 11-12 is rejected under 35 U.S.C. 103(a) as being unpatentable over Okuno ("Separating three simultaneous speeches with two microphones by integrating auditory and visual processing, Eurospeech 2001 – Scandinavia) in view of Sakagami (US 6,853,880), and in further view of Spors (Joint audio–video object localization and tracking, IEEE Signal Process. Mag. 18(1), 22–31, 2001).

4. **Regarding claims 1, 4 and 5**, Okuno discloses a robotics visual and auditory system including:

an auditory module for collecting external sounds by at least a pair of microphones, and, determining a direction of at least one speaker by sound source separation and localization by grouping based on pitch extraction and harmonic sounds

Art Unit: 3667

from the sound signals collected by the microphones, and extracting an auditory event (Introduction and Fig. 1);

a stereo module for extracting and localizing a longitudinally long matter based on a parallax extracted from images taken by a stereo camera and extracting a stereo event (2.1 Stereo Visual Processing);

a speech recognition part for conducting speech recognitions (Abstract), including:

a plurality of acoustic models; a speech recognition engine for processing a plurality of separated sound signals from respective sound sources to execute speech recognition processes by using the acoustic models (Fig. 1, see "Matching" and section 3.1 Benchmark Sounds, teaching use of HMM, which are acoustic models), and

a selector for integrating a plurality of speech recognition process results obtained by the speech recognition engine and selecting any one of speech recognition process results (Fig. 1 and equation 4);

in order to respond to cases the case where a plurality of speakers speak to the robot from different directions with respect to a robot's front direction as a base, the acoustic models are provided with respect to each speaker and each direction (Fig. 1, section 2 Direction-Pass Filter, see step 3; and section 3.3);

the auditory module collects sub-bands having interaural phase difference (IPD) or interaural intensity difference (IID) within a predetermined range by an active direction pass filter having a pass range which, becomes minimum in a frontal direction and becomes larger as an angle becomes wider to left and right on the basis of an

Art Unit: 3667

accurate sound source directional information from the association module and conducts sound source separation by restructuring a wave shape of a sound source (Fig. 1 and section 2.2);

the speech recognition engine conducts speech recognition using a plurality of the acoustic models in parallel for one sound signal separated by sound source separation (Fig. 1); and

the selector integrates speech recognition results from each acoustic model and judges a most reliable speech recognition result among the speech recognition results (Fig. 1 and equation 4).

Okuno discloses taking images of a robot's front by camera, however Okuno does not explicitly detail that the face module identifies each speaker. Sakagami discloses a similar robot that identifies each speaker, and extracts a face event from each speaker's face recognition and localization, based on images taken by the camera (Abstract, Figs. 10A and 10B). Sakagami further discloses:

a motor control module for rotating a robot in a horizontal direction by a drive motor and extracting a motor event based on a rotational position of the drive motor (Abstract); and

an attention control module for conducting an attention control based on the association stream, the auditory stream, the face stream and the stereo visual stream, and controlling the motor based on an action planning results accompanying the attention control (col. 2, lines 50-63).

Sakagami teaches that these features are useful for controlling the robot to behave in a more human-like manner (col. 2, lines 50-63). It would have been obvious to one of ordinary skill in the art at the time of the invention to use these features disclosed by Sakagami with the system disclosed by Okuno in order to control the robot to behave in a more human-like manner.

Okuno in view of Sakagami does not explicitly teach an association module as described in the remaining claim limitation, however Spors discloses a system for joint audio-visual object tracking, including an association module for determining each speaker's direction based on directional information of sound source localization of the auditory event and face localization of the face event from the auditory, face, stereo, and motor events, generating an auditory stream, a face stream and a stereo visual stream by respectively connecting auditory events, face events, and stereo events in a temporal direction using a Kalman filter for determinations, and further generating an association stream associating the auditory stream with the face and stereo visual streams (section 4, p. 395). Spors teaches that these features are useful for increasing the robustness of object localization and tracking algorithms. It would have been obvious to one of ordinary skill in the art at the time of the invention to use these features disclosed by Spors with the system disclosed by Okuno in view of Sakagami in order to provide a more robust tracking and recognition system for the robot.

5. **Regarding claims 2 and 12**, Okuno further discloses that the selector calculates a cost function value based on the recognition result by the speech recognition process and the speaker's direction upon integrating the speech recognition process result, and

Art Unit: 3667

judges a speech recognition process result having the maximum value of a cost function as a most reliable speech recognition result (section 2.2).

6. **Regarding claim 6**, Okuno does not explicitly teach the attention control module, however Sakagami further discloses that the attention control module is made up so as to collect speeches again from the microphones after the microphones turn to the sound source direction of the sound signals, and to perform again speech recognition of the speech by the auditory module based on the sound signals conducted sound source localization and sound source separation (col. 9, lines 13-63 and Figs. 2 and 8, after determining that sound is human voice, robot turns head toward sound and speech recognition process repeats). It would have been obvious to one of ordinary skill in the art at the time of the invention to use these features disclosed by Sakagami with the system disclosed by Okuno in order to control the robot to behave in a more human-like manner.

7. **Regarding claim 7**, Okuno does not explicitly teach the face module, however Sakagami further discloses that the auditory module refers to the face event from the face module upon performing the speech recognition (col. 9, lines 46-col. 10, line 37, face event is used to interpret sound command to robot). It would have been obvious to one of ordinary skill in the art at the time of the invention to use these features disclosed by Sakagami with the system disclosed by Okuno in order to control the robot to behave in a more human-like manner.

8. **Regarding claim 8**, Okuno further discloses that the auditory module refers to the stereo event from the stereo module upon performing the speech recognition (Fig. 1).

9. **Regarding claim 9**, Okuno further discloses that the auditory module refers to the stereo event from the stereo module upon performing the speech recognition (Fig. 1). Okuno does not explicitly teach the face module, however Sakagami further discloses that the auditory module refers to the face event from the face module upon performing the speech recognition (col. 9, lines 46-col. 10, line 37, face event is used to interpret sound command to robot). It would have been obvious to one of ordinary skill in the art at the time of the invention to use these features disclosed by Sakagami with the system disclosed by Okuno in order to control the robot to behave in a more human-like manner.

10. **Regarding claim 11**, Okuno further discloses that a pass range of the active direction pass filter can be controlled for each frequency. (Fig. 1).

11. Claims 3 and 10 are rejected under 35 U.S.C. 103(a) as being unpatentable over Okuno ("Separating three simultaneous speeches with two microphones by integrating auditory and visual processing, Eurospeech 2001 – Scandinavia) in view of Sakagami (US 6,853,880), in further view of Spors (Joint audio–video object localization and tracking, IEEE Signal Process. Mag. 18(1), 22–31, 2001) and in further view of Shimomura (US 2001/0021909).

12. **Regarding claims 3 and 10**, Okuno in view of Sakagami and Spors does not explicitly detail a dialogue part, however Shimomura further discloses a similar robot that is provided with a dialogue part to output the speech recognition process results selected by the selector to outside (Fig. 3, speech synthesizer 36; see also [0064]). Shimomura teaches that these features are useful for controlling the robot to hold a conversation with a person (see [0002]). It would have been obvious to one of ordinary skill in the art at the time of the invention to use these features disclosed by Shimomura with the system disclosed by Okuno in view of Sakagami and Spors in order to control the robot to hold a conversation with a person.

13. Claim 13 is rejected under 35 U.S.C. 103(a) as being unpatentable over Okuno ("Separating three simultaneous speeches with two microphones by integrating auditory and visual processing, Eurospeech 2001 – Scandinavia) in view of Sakagami (US 6,853,880), in further view of Spors (Joint audio–video object localization and tracking, IEEE Signal Process. Mag. 18(1), 22–31, 2001) and in further view of Bancroft (US 2002/0165638).

14. **Regarding claim 13**, Okuno in view of Sakagami and Spors does not explicitly detail name recognition, however Bancroft discloses a robot for use in retail environments and teaches that it is known that a robot system may utilize voice recognition techniques to identify a particular person (see [0136]). Bancroft teaches that these features allow the robot to more effectively interact with a person. It would have been obvious to one of ordinary skill in the art at the time of the invention to use

Art Unit: 3667

these features disclosed by Bancroft with the system disclosed by Okuno in view of Sakagami and Spors in order to allow the robot to more effectively interact with a person.

Conclusion

15. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to JERRAH EDWARDS whose telephone number is (571) 270-3044. The examiner can normally be reached on Monday through Friday, 10:00 AM - 6:30 PM.

Art Unit: 3667

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, James Trammell can be reached on 571-272-6712. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

/J. E./
Examiner, Art Unit 3667

/Mary Cheung/
Primary Examiner, Art Unit 3667